Structure-Blind Signal Recovery

Dmitry Ostrovsky, Zaid Harchaoui, Anatoli B. Juditsky, Arkadi Nemirovski

Structure-Blind Signal Recovery

Dmitry Ostrovsky* Zaid Harchaoui† Anatoli Juditsky‡ Arkadi Nemirovski§

July 17, 2016

Abstract

We consider the problem of recovering a signal observed in Gaussian noise. If the set of signals is convex and compact, and can be specified beforehand, one can use classical linear estimators that achieve a risk within a constant factor of the minimax risk [5]. However, when the set is unspecified, designing an estimator that is blind to the hidden structure of the signal remains a challenging problem. We propose a new family of estimators to recover signals observed in Gaussian noise. Instead of specifying the set where the signal lives, we assume the existence of a well-performing linear estimator. Proposed estimators enjoy exact oracle inequalities and can be efficiently computed through convex optimization. We present several numerical illustrations that show the potential of the approach.

1 Introduction

We consider the problem of recovering a *complex-valued signal* $(x_t)_{t\in\mathbb{Z}}$ from the noisy observations

$$y_\tau = x_\tau + \sigma\zeta_\tau, \quad -n \le \tau \le n. \tag{1}$$

Here $n \in \mathbb{Z}_+$, and $\zeta_\tau \sim \mathbb{C}\mathcal{N}(0,1)$ are i.i.d. standard complex-valued Gaussian random variables, meaning that $\zeta_0 = \xi_0^1 + \imath\xi_0^2$ with i.i.d. $\xi_0^1, \xi_0^2 \sim \mathcal{N}(0,1)$. Our goal is to recover x_t given the sequence of observations $y_{-n}, ..., y_t$ up to instant t, a task usually referred to as (pointwise) *filtering* in machine learning, statistics, and signal processing [8].

The traditional approach to this problem considers *linear estimators*, or linear filters, which write as

$$\widehat{x}_t = \sum_{\tau=0}^{t+n+1} \phi_\tau y_{t-\tau}, \quad -n \le t \le n.$$

Linear estimators have been thoroughly studied in various forms, they are both theoretically attractive [20, 10, 6, 5] and easy to use in practice. If the set $\mathcal{X}$ of signals is well-specified, one can usually compute a (nearly) minimax on $\mathcal{X}$ linear estimator in a closed form. In particular, if $\mathcal{X}$ is a class of smooth signals, such as a Hölder or a Sobolev ball, then the corresponding estimator is given by the kernel estimator with the properly set bandwidth parameter [20] and is minimax among all possible estimators. Moreover, as shown by [9, 5], if only $\mathcal{X}$ is convex, compact, and centrally symmetric, the risk of the best linear estimator of x_t is within a

*LJK, University of Grenoble Alpes, 700 Avenue Centrale 38401 Domaine Universitaire de Saint-Martin-d'Hres, France, `dmitrii.ostrovskii@imag.fr`

†Courant Institute for Mathematical Sciences, New York University, 715 Broadway, New York, NY 10003, USA, `zaid.harchaoui@nyu.edu`

‡LJK, University of Grenoble Alpes, 700 Avenue Centrale 38401 Domaine Universitaire de Saint-Martin-d'Hres, France, `juditsky@imag.fr`

§Georgia Institute of Technology, Atlanta, Georgia 30332, USA, `nemirovs@isye.gatech.edu`

The first, second and third authors were supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025) and the project Titan (CNRS-Mastodons). The second author was also supported by the project Macaron (ANR-14-CE23-0003-01) and the program "Learning in Machines and Brains" (CIFAR). Research of the fourth author was supported by NSF grants CMMI-1262063, CCF-1523768.

small constant factor of the minimax risk over $\mathcal{X}$. Besides, if the set $\mathcal{X}$ can be specified in a computationally tractable way, which clearly is still a weaker assumption than classical smoothness assumptions, the best linear estimator can be efficiently computed by solving a convex optimization problem on $\mathcal{X}$. In other words, given a computationally tractable set $\mathcal{X}$ on the input, one can *compute* a nearly-minimax linear estimator and the corresponding (nearly-minimax) risk over $\mathcal{X}$. The strength of this approach, however, comes at a price: the set $\mathcal{X}$ still must be specified beforehand. Therefore, when one faces a recovery problem *without any prior knowledge of* $\mathcal{X}$, this approach cannot be implemented.

We adopt here a novel approach to filtering, which we refer to as *structure-blind recovery*. While we do not require $\mathcal{X}$ to be specified beforehand, we assume that there exists a *linear oracle* – a well-performing linear estimator of x_t. Previous works [12, 7], following a similar philosophy, proved that one can efficiently adapt to the linear oracle filter of length $m = O(n)$ if the corresponding filter ϕ is *time-invariant*, i.e. it recovers the target signal uniformly well in the $O(n)$-sized neighbourhood of t, and if its ℓ_2-norm is small – bounded by $\rho/\sqrt{m}$ for a moderate $\rho \geq 1$. The adaptive estimator is computed by minimizing the ℓ_∞-norm of the filter discrepancy, in the Fourier domain, under the constraint on the ℓ_1-norm of the filter in the Fourier domain. Put in contrast to the oracle linear filter, the price for adaptation is proved to be of order $O(\rho^3\sqrt{\ln n})$, with the lower bound of $O(\rho\sqrt{\ln n})$ [12, 7].

We make the following contributions:

- we propose a new family of recovery methods, obtained by solving a least-squares problem constrained or penalized by the ℓ_1-norm of the filter in the Fourier domain;
- we prove exact oracle inequalities for the ℓ_2-risk of these methods;
- we show that the price for adaptation improves upon previous works [12, 7] to $O(\rho^2\sqrt{\ln n})$ for the point-wise risk and to $O(\rho\sqrt{\ln n})$ for the ℓ_2-risk.
- we present numerical experiments that show the potential of the approach on synthetic and real-world images and signals.

Before presenting the theoretical results, let us introduce the notation we use throughout the paper.

Filters. Let $\mathbb{C}(\mathbb{Z})$ be the linear space of all two-sided complex-valued sequences $x = \{x_t \in \mathbb{C}\}_{t\in\mathbb{Z}}$. For $k, k' \in \mathbb{Z}$ we consider finite-dimensional subspaces

$$\mathbb{C}(\mathbb{Z}_k^{k'}) = \{x \in \mathbb{C}(\mathbb{Z}) : \quad x_t = 0, \quad t \notin [k, k']\}\,.$$

It is convenient to identify m-dimensional complex vectors, $m = k' - k + 1$, with elements of $\mathbb{C}(\mathbb{Z}_k^{k'})$ by means of the notation:

$$x_k^{k'} := [x_k; \ldots; x_{k'}] \in \mathbb{C}^{k'-k+1}.$$

We associate to linear mappings $\mathbb{C}(\mathbb{Z}_k^{k'}) \to \mathbb{C}(\mathbb{Z}_j^{j'})$ $(j' - j + 1) \times (k' - k + 1)$ matrices with complex entries. The *convolution* $u * v$ of two sequences $u, v \in \mathbb{C}(\mathbb{Z})$ is a sequence with elements

$$[u * v]_t = \sum_{\tau\in\mathbb{Z}} u_\tau v_{t-\tau}, \quad t \in \mathbb{Z}.$$

Given observations (1) and $\varphi \in \mathbb{C}(\mathbb{Z}_0^m)$ consider the *(left) linear estimation* of x associated with *filter* φ:

$$\widehat{x}_t = [\varphi * y]_t$$

($\widehat{x}_t$ is merely a kernel estimate of x_t by a kernel φ supported on $[0, \ldots, m]$). Note also that if Δ is the right-shift operator on $\mathbb{C}(\mathbb{Z})$, $[\Delta x]_t = x_{t-1}$, the linear estimation $[\varphi * y]_t$ may be alternatively written as $[\varphi(\Delta)y]_t$.

Discrete Fourier transform. We define the unitary *Discrete Fourier transform* (DFT) operator $F_n : \mathbb{C}^{n+1} \to \mathbb{C}^{n+1}$ by

$$z \mapsto F_n z, \quad [F_n z]_k = (n+1)^{-1/2} \sum_{t=0}^{n} z_t \, e^{\frac{2\pi \imath k t}{n+1}}, \quad 0 \le k \le n.$$

The *inverse Discrete Fourier transform* (iDFT) operator F_n^{-1} is given by $F_n^{-1} := F_n^H$ (here A^H stands for Hermitian adjoint of A). By the Fourier inversion theorem, $F_n^{-1}(F_n z) = z$.

We denote $\|\cdot\|_p$ usual ℓ_p-norms on $\mathbb{C}(\mathbb{Z})$: $\|x\|_p = (\sum_{t\in\mathbb{Z}} |x_t|^p)^{1/p}$, $p \in [1,\infty]$. Usually, the argument will be finite-dimensional – an element of $\mathbb{C}(\mathbb{Z}_k^{k'})$; we reserve the special notation

$$\|x\|_{n,p} := \|x_0^n\|_p.$$

Furthermore, DFT allows to equip $\mathbb{C}(\mathbb{Z}_0^n)$ with the norms associated with ℓ_p-norms in the spectral domain:

$$\|x\|_{n,p}^* := \|x_0^n\|_p^* := \|F_n x_0^n\|_p, \quad p \in [1,\infty];$$

note that unitarity of the DFT implies the Parseval identity: $\|x\|_{n,2} = \|x\|_{n,2}^*$.

Finally, c, C, and C' stand for generic absolute constants.

2 Oracle inequalities for adaptive recovery procedures

2.1 Constrained recovery

Given observations (1) and $\overline{\varrho} > 0$, we first consider the *constrained recovery* $\widehat{x}_{\rm con}$ given by

$$[\widehat{x}_{\rm con}]_t = [\widehat{\varphi} * y]_t, \quad t = 0, ..., n,$$

where $\widehat{\varphi}$ is an optimal solution of the constrained optimization problem

$$\min_{\varphi \in \mathbb{C}(\mathbb{Z}_0^n)} \left\{ \|y - \varphi * y\|_{n,2} : \ \|\varphi\|_{n,1}^* \le \overline{\varrho}/\sqrt{n+1} \right\}. \tag{2}$$

The constrained recovery estimator minimizes a least-squares fit criterion under a constraint on $\|\varphi\|_{n,1}^* = \|F_n \varphi_0^n\|_1$, that is an ℓ_1 constraint on the discrete Fourier transform of the filter. While the least-squares objective naturally follows from the Gaussian noise assumption, the constraint can be motivated as follows.

Small-error linear filters. Linear filter φ^o with a small ℓ_1 norm in the spectral domain and small recovery error exists, essentially, whenever there exists a linear filter with small recovery error [12, 7]. Indeed, let us say that $x \in \mathbb{C}(\mathbb{Z}_0^n)$ is *simple* [7] *with parameters* $m \in \mathbb{Z}_+$ *and* $\rho \ge 1$ if there exists $\phi^o \in \mathbb{C}(\mathbb{Z}_0^m)$ such that for all $-m \le \tau \le 2m$,

$$\left[\mathbf{E}\left\{|x_\tau - [\phi^o * y]_\tau|^2\right\}\right]^{1/2} \le \frac{\sigma\rho}{\sqrt{m+1}}. \tag{3}$$

In other words, x is (m,ρ)-simple if there exists a hypothetical filter ϕ^o of the length at most $m+1$ which recovers x_τ with squared risk uniformly bounded by $\frac{\sigma^2\rho^2}{m+1}$ in the interval $-m \le \tau \le 2m$. Note that (3) clearly implies that $\|\phi^o\|_2 \le \rho/\sqrt{m+1}$, and that $|[x - \phi^o * x]_\tau| \le \sigma\rho/\sqrt{m+1}$ $\forall \tau$, $-m \le \tau \le 2m$. Now, let $n = 2m$, and let

$$\varphi^o = \phi^o * \phi^o \in \mathbb{C}^{n+1}.$$

As proved in Appendix C, we have

$$\|\varphi^o\|_{n,1}^* \le 2\rho^2/\sqrt{n+1}, \tag{4}$$

and, for a moderate absolute constant c,

$$\|x - \varphi^o * y\|_{n,2} \le c\sigma\rho^2\sqrt{1 + \ln[1/\alpha]} \tag{5}$$

with probability $1-\alpha$. To summarize, *if x is (m,ρ)-simple, i.e., when there exists a filter ϕ^o of length $\le m+1$ which recovers x with small risk on the interval $[-m, 2m]$, then the filter $\varphi^o = \phi^o * \phi^o$ of the length at most $n+1$, with $n = 2m$, has small norm $\|\varphi^o\|_{n,1}^*$ and recovers the signal x with (essentially the same) small risk on the interval $[0, n]$.*

Hidden structure. The constrained recovery estimator is completely blind to a possible hidden structure of the signal, yet can seamlessly adapt to it when such a structure exists, in a way that we can rigorously establish. We formalize the hidden structure as an unknown shift-invariant linear subspace of $\mathbb{C}(\mathbb{Z})$, $\Delta\mathcal{S} = \mathcal{S}$, of a small dimension s. We do not assume that x belongs to that subspace. Instead, we make a more general assumption that x is *close* to this subspace, that is, it may be decomposed into a sum of a component that lies in the subspace and a component whose norm we can control.

Assumption A *We suppose that x admits the decomposition*

$$x = x^{\mathcal{S}} + \varepsilon, \quad x^{\mathcal{S}} \in \mathcal{S},$$

where $\mathcal{S}$ is (an unknown) shift-invariant, $\Delta\mathcal{S} = \mathcal{S}$, subspace of $\mathbb{C}(\mathbb{Z})$ of dimension s, $1 \le s \le n+1$, and ε is "small", namely,

$$\|\Delta^\tau \varepsilon\|_{n,2} \le \sigma\varkappa, \quad 0 \le \tau \le n.$$

Examples. Shift-invariant subspaces of sequences $\mathbb{C}(\mathbb{Z})$ are actually sets of solutions of difference equations

$$\sum_{i=0}^{s} p_i x_{\tau-i} = 0, \;\; \tau \in \mathbb{Z}. \tag{6}$$

For instance, *discrete-time polynomials* $x_\tau = \sum_{i=0}^s a_i \tau^i$, $\tau \in \mathbb{Z}$ of degree s form a linear space of dimension $s+1$ of solutions of the equation (6) with binomial coefficients $p_i = (-1)^i \binom{n}{i}$. Harmonic oscillations are another example: $x_\tau = Ce^{\imath\omega\tau}$ for some frequency $\omega \in [0, 2\pi)$ is the set of solutions of the equation $x_\tau - e^{\imath\omega} x_{\tau-1} = 0$.

We can now state an oracle inequality for the constrained recovery estimator; see Appendix B for the proof.

Theorem 2.1. *Let $\overline{\varrho} \ge 1$, and let $\varphi^o \in \mathbb{C}(\mathbb{Z}_0^n)$ be such that*

$$\|\varphi^o\|_{n,1}^* \le \overline{\varrho}/\sqrt{n+1}.$$

Suppose that Assumption A holds for some $s \in \mathbb{Z}_+$ and $\varkappa < \infty$. Then for any α, $0 < \alpha \le 1$, it holds with probability at least $1-\alpha$:

$$\|x - \widehat{x}_{\mathrm{con}}\|_{n,2} \le \|x - \varphi^o * y\|_{n,2} + C\sigma\sqrt{s + \overline{\varrho}\left(\varkappa\sqrt{\ln\left[1/\alpha\right]} + \ln\left[n/\alpha\right]\right)}. \tag{7}$$

When considering *simple* signals, Theorem 2.1 gives the following.

Corollary 2.1. *Assume that signal x is (m,ρ)-simple, $\rho \ge 1$ and $m \in \mathbb{Z}_+$. Let $n = 2m$, $\overline{\varrho} \ge 2\rho^2$, and let Assumption A hold for some $s \in \mathbb{Z}_+$ and $\varkappa < \infty$. Then for any $\alpha, 0 < \alpha \le 1$, it holds with probability at least $1-\alpha$:*

$$\|x - \widehat{x}_{\mathrm{con}}\|_{n,2} \le C\sigma\rho^2\sqrt{\ln[1/\alpha]} + C'\sigma\sqrt{s + \overline{\varrho}\left(\varkappa\sqrt{\ln\left[1/\alpha\right]} + \ln\left[n/\alpha\right]\right)}.$$

Adaptation and price. The price for adaptation in Theorem 2.1 and Corollary 2.1 is determined by three parameters: the bound on the filter norm $\overline{\varrho}$, the deterministic error $\varkappa$, and the subspace dimension s. Assuming that the signal to recover is simple, and that $\overline{\varrho} = 2\rho^2$, let us compare the magnitude of the oracle error to the term of the risk which reflects "price of adaptation". Typically (in fact, in all known to us cases of recovery of signals from a shift-invariant subspace), the parameter ρ is at least $\sqrt{s}$. Therefore, the bound (5) implies the "typical bound" $O(\sigma\sqrt{\gamma}\rho^2) = \sigma s\sqrt{\gamma}$ for the term $\|x - \varphi^o * y\|_{n,2}$ (we denote $\gamma = \ln(1/\alpha)$). As a result, for instance, in the "parametric situation", when the signal belongs or is very close to the subspace, that is when $\varkappa = O(\ln(n))$, the price of adaptation $O\left(\sigma[s + \rho^2(\gamma + \sqrt{\gamma}\ln n)]^{1/2}\right)$ is much smaller than the bound on the oracle error. In the "nonparametric situation", when $\varkappa = O(\rho^2)$, the price of adaptation has same order of magnitude as the oracle error.

Finally, note that under the premise of Corollary 2.1 we can also bound the pointwise error; see Appendix B for the proof. We state the result for $\overline{\varrho} = 2\rho^2$ for simplicity.

Theorem 2.2. *Assume that signal x is (m,ρ)-simple, $\rho \ge 1$ and $m \in \mathbb{Z}_+$. Let $n = 2m$, $\overline{\varrho} = 2\rho^2$, and let Assumption A hold for some $s \in \mathbb{Z}_+$ and $\varkappa < \infty$. Then for any $\alpha, 0 < \alpha \le 1$, the constrained recovery $\widehat{x}_{\rm con}$ satisfies*

$$|x_n - [\widehat{x}_{\rm con}]_n| \le C\frac{\sigma\rho}{\sqrt{m+1}}\left[\rho^2\sqrt{\ln[n/\alpha]} + \rho\sqrt{\varkappa\sqrt{\ln[1/\alpha]}} + \sqrt{s}\right].$$

2.2 Penalized recovery

The constrained recovery estimator assumes that the parameter $\overline{\varrho}$ is known. If the noise variance is known (or can be estimated from data), then we can build a more practical estimator that still enjoys an oracle inequality.

The *penalized recovery* estimator $[\widehat{x}_{\rm pen}]_t = [\widehat{\varphi} * y]_t$ is an optimal solution to a regularized least-squares minimization problem, where the regularization penalizes the ℓ_1-norm of the filter in the Fourier domain:

$$\widehat{\varphi} \in \underset{\varphi \in \mathbb{C}(\mathbb{Z}_0^n)}{\mathrm{Argmin}} \left\{ \|y - \varphi * y\|_{n,2}^2 + \lambda\sqrt{n+1}\,\|\varphi\|_{n,1}^* \right\}. \tag{8}$$

We establish an oracle inequality for the penalized recovery estimator, similarly to Theorem 2.1:

Theorem 2.3. *Let Assumption A hold for some $s \in \mathbb{Z}_+$ and $\varkappa < \infty$, and let $\varphi^o \in \mathbb{C}(\mathbb{Z}_0^n)$ satisfy $\|\varphi^o\|_{n,1}^* \le \varrho/\sqrt{n+1}$ for some $\varrho \ge 1$.*

1^o**.** *Suppose that the regularization parameter of penalized recovery $\widehat{x}_{\rm pen}$ satisfies $\lambda \ge \underline{\lambda}$,*

$$\underline{\lambda} := 60\sigma^2 \ln[63n/\alpha].$$

Then, for $0 < \alpha \le 1$, it holds with probability at least $1-\alpha$:

$$\|x - \widehat{x}_{\rm pen}\|_{n,2} \le \|x - \varphi^o * y\|_{n,2} + C\sqrt{\varrho\lambda} + C'\sigma\sqrt{s + (\widehat{\varrho}+1)\varkappa\sqrt{\ln[1/\alpha]}},$$

where $\widehat{\varrho} := \sqrt{n+1}\,\|\widehat{\varphi}\|_{n,1}^*$.

2^o**.** *Moreover, if $\varkappa \le \bar{\varkappa}$,*

$$\bar{\varkappa} := \frac{10\ln[42n/\alpha]}{\sqrt{\ln[16/\alpha]}},$$

and $\lambda \ge 2\underline{\lambda}$, one has

$$\|x - \widehat{x}_{\rm pen}\|_{n,2} \le \|x - \varphi^o * y\|_{n,2} + C\sqrt{\varrho\lambda} + C'\sigma\sqrt{s}.$$

The proof of Theorem 2.3 closely follows that of Theorem 2.1 and is omitted.

2.3 Discussion

There is some redundancy between "simplicity" of a signal, as defined by (3), and Assumption A. Usually a simple signal or image x is also close to a low-dimensional subspace of $\mathbb{C}(\mathbb{Z})$ (see, e.g., [13, section 4]), so that Assumption A holds "automatically". Likewise, x is "almost" simple when it is close to a low-dimensional time-invariant subspace. Indeed, if $x \in \mathbb{C}(\mathbb{Z})$ *belongs* to $\mathcal{S}$, i.e. Assumption A holds with $\varkappa = 0$, one can easily verify that for $n \ge s$ there exists a filter $\phi^o \in \mathbb{C}(\mathbb{Z}_{-n}^n)$ such that

$$\|\phi^o\|_2 \le \sqrt{s/(n+1)}, \text{ and } x_\tau = [\phi^o * x]_\tau,\ \ \tau \in \mathbb{Z}. \tag{9}$$

See Appendix C for the proof. This implies that x can be recovered efficiently from observations (1):

$$\left[\mathbf{E}\left\{|x_\tau - [\phi^o * y]_\tau|^2\right\}\right]^{1/2} \leq \sigma\sqrt{\frac{s}{n+1}}.$$

In other words, if instead of the filtering problem we were interested in the *interpolation* problem of recovering x_t given $2n+1$ observations $y_{t-n},...,y_{t+n}$ on the left *and* on the right of t, Assumption A would imply a kind of simplicity of x. On the other hand, it is clear that Assumption A is not sufficient to imply the simplicity of x "with respect to the filtering", in the sense of the definition we use in this paper, when we are allowed to use only observations on the left of t to compute the estimation of x_t. Indeed, one can see, for instance, that already signals from the parametric family $\mathcal{X}_\alpha = \{x \in \mathbb{C}(\mathbb{Z}) : x_\tau = c\alpha^\tau,\, c \in \mathbb{C}\}$, with a given $|\alpha| > 1$, which form a one-dimensional space of solutions of the equation $x_\tau = \alpha x_{\tau-1}$, cannot be estimated with small risk at t using only observations on the left of t (unless $c = 0$), and thus are not simple in the sense of (3).

Of course, in the above example, the "difficulty" of the family $\mathcal{X}_\alpha$ is due to instability of solutions of the difference equation which explode when $\tau \to +\infty$. Note that signals $x \in \mathcal{X}_\alpha$ with $|\alpha| \leq 1$ (linear functions, oscillations, or damped oscillations) are simple. More generally, suppose that x satisfies a difference equation of degree s:

$$0 = p(\Delta)x_\tau \left[= \sum_{i=0}^{s} p_i x_{\tau-i}\right], \tag{10}$$

where $p(z) = \sum_{i=0}^s p_i z^i$ is the corresponding characteristic polynomial and Δ is the right shift operator. When $p(z)$ is unstable – has roots *inside* the unit circle – (depending on "initial conditions") the set of solutions to the equation (10) contains difficult to filter signals. Observe that stability of solutions is related to the direction of the time axis; when the characteristic polynomial $p(z)$ has roots *outside* the unit circle, the corresponding solutions may be "left unstable" – increase exponentially when $\tau \to -\infty$. In this case "right filtering" – estimating x_τ using observations on the right of τ – will be difficult. A special situation where interpolation and filtering is always simple arises when the characteristic polynomial of the difference equation has all its roots on the unit circle. In this case, solutions to (10) are "generalized harmonic oscillations" (harmonic oscillations modulated by polynomials), and such signals are known to be simple. Theorem 2.4 summarizes the properties of the solutions of (10) in this particular case; see Appendix C.1 for the proof.

Theorem 2.4. *Let s be a positive integer, and let $p = [p_0;...;p_s] \in \mathbb{C}^{s+1}$ be such that the polynomial $p(z) = \sum_{i=0}^s p_i z^i$ has all its roots on the unit circle. Then for every integer m satisfying*

$$m \geq m(s) := Cs^2\ln(s+1),$$

one can point out $q \in \mathbb{C}^{m+1}$ such that any solution to (10) satisfies

$$x_\tau = [q * x]_\tau, \quad \forall \tau \in \mathbb{Z},$$

and

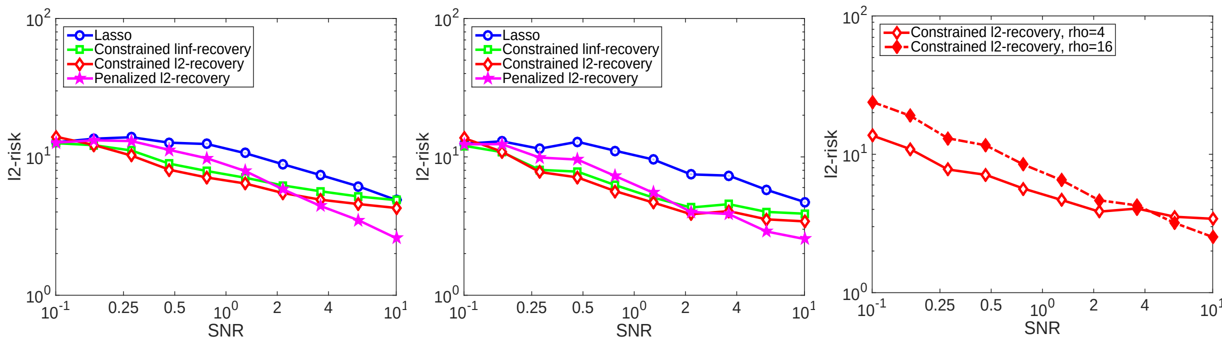
$$\|q\|_2 \leq \rho(s,m)/\sqrt{m} \quad \textit{where} \quad \rho(s,m) = C' \min\left\{s^{3/2}\sqrt{\ln s},\, s\sqrt{\ln[ms]}\right\}. \tag{11}$$

3 Numerical experiments

We present a preliminary simulation study of the proposed recovery algorithms in several application scenarios. We illustrate the performance of *constrained ℓ_2-recovery* of Sec. 2.1 and the *penalized ℓ_2-recovery* of Sec. 2.2 and compare it to that of the *constrained ℓ_∞-recovery* of [7] and the Lasso recovery of [1] for signal and image denoising. Some implementation details are discussed in Appendix D. Details and discussion of the discretization approach underlying the competing Lasso method can be found in Sec. 3.6 of [1]. First, we investigate the behavior of the different methods on a simple one-dimensional signal denoising example.

3.1 Signal denoising

We consider two different scenarios. In Scenario 1, the signal is a sum of $k = 4$ sine waves of equal amplitudes and random frequencies. In Scenario 2, $k/2 = 2$ pairs of close frequencies are sampled, with the frequencies in each pair separated by only 0.1 of the DFT bin $2\pi/n$, making recovery harder due to high signal coherency. The signal-to-noise ratio (SNR) is defined as the ratio $\|x\|_{n,2}^2/(\sigma^2 n)$. In both scenarios, we perform 100 Monte-Carlo trials for a signal-to-noise ratio in the range $10^{-1}, \dots, 10^1$. The parameter of the constrained ℓ_2- and constrained ℓ_∞-recoveries was set $\overline{\varrho} = 4$. We use the same regularization parameter of the penalized ℓ_2-recovery as in the dimension reduction example above. For the Lasso, we use the theoretically recommended value [1]. As Fig. 1 suggests, by comparing results with $\overline{\varrho} = 4$ and $\overline{\varrho} = 16$, the constrained ℓ_2-recovery method is in fact not too sensitive to the choice of $\overline{\varrho}(k)$, where k is the number of sines.



Scenario 1 | Scenario 2 | $\overline{\varrho} = 4$ vs. $\overline{\varrho} = 16$.

Figure 1: Denoising harmonic oscilations.

3.2 Image Denoising

We now consider the recovery of an unknown regression function f on the regular grid $\mathcal{T}$ on $[0,1]^2$ given noisy observations:

$$y_\tau = f(\tau) + \sigma\zeta_\tau, \quad \tau \in \mathcal{T} := \left\{\frac{1}{n}, \frac{2}{n}, ..., 1\right\}^2 . \tag{12}$$

The signal-to-noise ratio (SNR) is defined here as $\|x\|_{n,2}^2/(\sigma^2 n)$.

Denoising textures. In this experiment, we apply the proposed recovery methods to denoise two images from the original Brodatz texture database, observed in white Gaussian noise.

The signal-to-noise ratio (SNR) is set to SNR $= 1$. We use the blockwise implementation of the constrained ℓ_2-recovery algorithm, as described in Appendix D. We set the constraint parameter to $\overline{\varrho} = 4$, and we use the adaptation procedure of [7, Sec. 3.2] to define the estimation bandwidth. We use the Lasso method of [1] with $\lambda = 0.5\sigma\sqrt{2\log N}$, halving the theoretically recommended value to prevent edge oversmoothing. The resulting images are presented in Fig. 2. Despite comparable quality in the mean square sense, the two methods significantly differ in their local behaviour. In particular, constrained ℓ_2-recovery better restores the local signal features, whereas the Lasso tends to oversmooth, even with the chosen value of the regularization penalty.

Dimension reduction. The purpose of this experiment is to illustrate the performance of the penalized ℓ_2-recovery in the problem of estimating a function with a hidden structure. We consider the single-index model of the regression function f:

$$f(t) = g(\theta^T t), \quad g(\cdot) \in \mathcal{S}_\beta^1(L), \quad \|\theta\|_1 = 1. \tag{13}$$

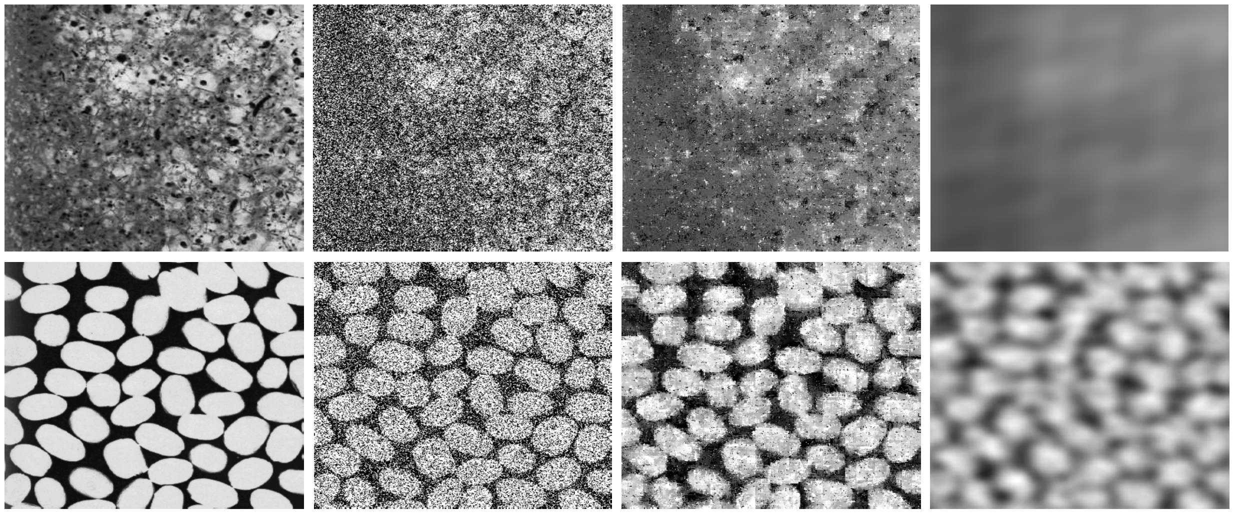


Figure 2: Recovery of two instances of the Original Brodatz database, cut by half to 320×320 and observed with SNR = 1. ℓ_2-error: $1.35e4$ for the constrained ℓ_2-recovery $1.25e4$ for the Lasso in the first row (inst. D73); $1.97e4$ for the constrained ℓ_2-recovery, $2.02e4$ for the Lasso method in the second row (inst. D75).

Here, $\mathcal{S}^1_\beta(L) = \{g : \mathbb{R} \to \mathbb{R}, \|g^{(\beta)}(\cdot)\|_2 \leq L\}$ is the Sobolev ball of smooth periodic functions on $[0, 1]$, and θ, called the index, is some unknown direction.

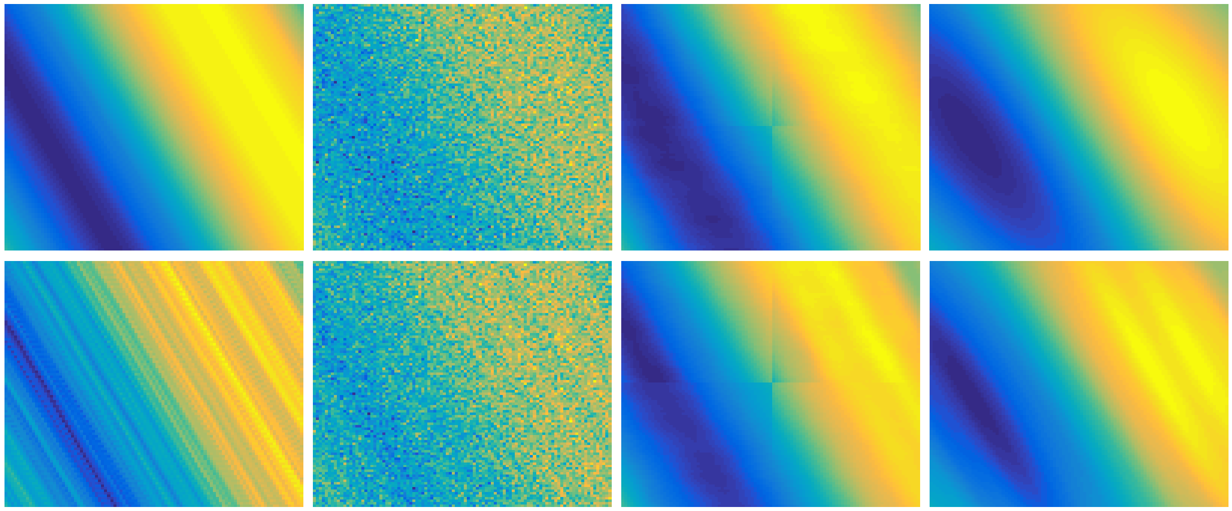
Note that if it is known a priori that the regression function possesses the structure (13), but the index is unknown, one can adapt to it, obtaining the one-dimensional rates of recovery; see e.g. [16] and reference therein.

β	Observations	Penalized recovery	Lasso
0.5	45.30	54.22	54.94
1.0	64.71	17.05	16.94
2.0	65.66	4.02	7.49
3.0	65.63	2.72	6.95

Table 1: ℓ_2-error in the dimension reduction experiment.

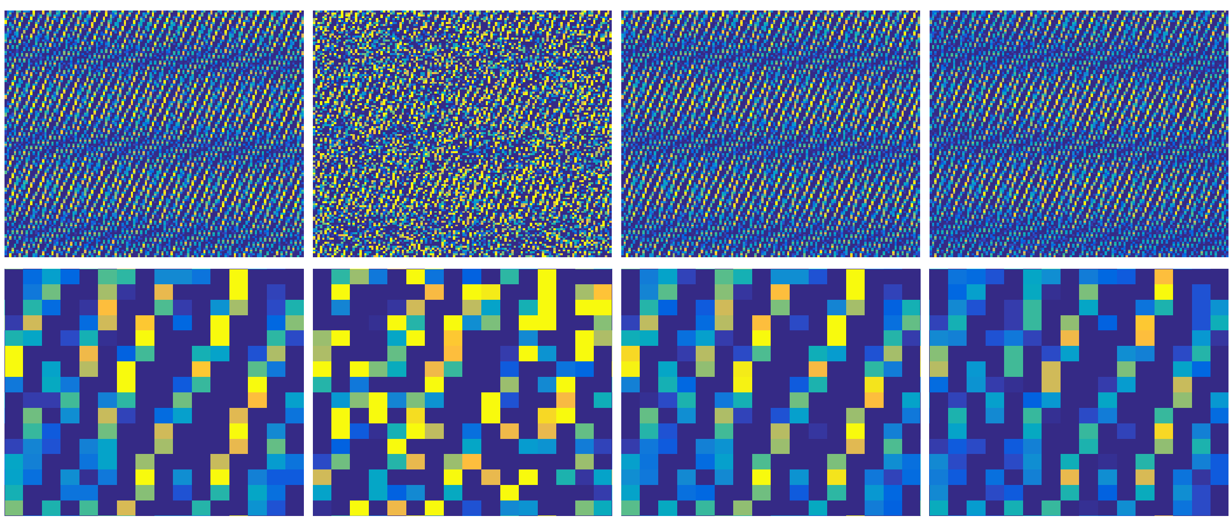
In our experiments, a random direction θ is sampled uniformly from the unit sphere and then renormalized. We use the blockwise strategy, with one block corresponding to the entire image. For penalized ℓ_2-recovery (8), the regularization penalty parameter is set to $\lambda = 2\sigma^2 \log(21/\alpha)$ with a confidence level $1 - \alpha = 0.9$. For Lasso, we use the same parameter setting as previously. The signal-to-noise ratio (SNR) is set to SNR = 1. The corresponding results are presented in Tab. 1 and Fig. 3.

Denoising harmonic oscillations. We consider the recovery of noisy images which are sums of $k = 4$ harmonic oscillations in $\mathbb{R}^2$ with random frequencies. We compare constrained ℓ_2-recovery (2), with a single block and the constraint parameter set to $\overline{\varrho}(k) = k^2$, to the Lasso, with the regularization penalty parameter set to the theoretically recommended value [1]. The signal-to-noise ratio (SNR) is set to SNR = 0.25. We present the results in Fig. 4.



True signal | Observations | Penalized recovery | Lasso

Figure 3: Recovery of the single-index signal (13), observed with SNR = 1, for $\beta = 2$ (first row) and $\beta = 1$ (second row). Results are in Tab. 1.



True signal | Observations | Constrained recovery | Lasso

Figure 4: Recovery of a sum of 4 random sines observed with SNR = 0.25. Second row: the magnified north-west corner of the image. ℓ_2-error: 6.61 for the constrained ℓ_2-recovery, 13.67 for the Lasso method.

References

[1] B. Bhaskar, G. Tang, and B. Recht. Atomic norm denoising with applications to line spectral estimation. *Signal Processing, IEEE Transactions on*, 61(23):5987–5999, 2013.

[2] P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, pages 1705–1732, 2009.

[3] E. Candes and T. Tao. The Dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics*, pages 2313–2351, 2007.

[4] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[5] D. L. Donoho. Statistical estimation and optimal recovery. *Ann. Statist.*, 22(1):238–270, 03 1994.

[6] D. L. Donoho and M. G. Low. Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.*, 20(2):944–970, 06 1992.

[7] Z. Harchaoui, A. Juditsky, A. Nemirovski, and D. Ostrovsky. Adaptive recovery of signals by convex optimization. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, pages 929–955, 2015.

[8] S. Haykin. *Adaptive filter theory.* Prentice-Hall, Inc., 1991.

[9] I. Ibragimov and R. Khasminskii. Nonparametric estimation of the value of a linear functional in gaussian white noise. *Theor. Probab. & Appl.*, 29:1–32, 1984.

[10] I. A. Ibragimov and R. Z. Khas' minskii. Estimation of linear functionals in gaussian noise. *Theory of Probability & Its Applications*, 32(1):30–39, 1988.

[11] A. Juditsky and A. Nemirovski. Functional aggregation for nonparametric regression. *Annals of Statistics*, pages 681–712, 2000.

[12] A. Juditsky and A. Nemirovski. Nonparametric denoising of signals with unknown local structure, I: Oracle inequalities. *Applied and Computational Harmonic Analysis*, 27(2):157–179, 2009.

[13] A. Juditsky and A. Nemirovski. Nonparametric denoising signals of unknown local structure, II: Nonparametric function recovery. *Applied and Computational Harmonic Analysis*, 29(3):354–367, 2010.

[14] A. Juditsky and A. Nemirovski. On detecting harmonic oscillations. *Bernoulli Journal*, 2013.

[15] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, pages 1302–1338, 2000.

[16] O. Lepski and N. Serdyukova. Adaptive estimation under single-index constraint in a regression model. *The Annals of Statistics*, 42(1):1–28, 2014.

[17] O. V. Lepskii. On a problem of adaptive estimation in gaussian white noise. *Theory of Probability & Its Applications*, 35(3):454–466, 1991.

[18] Y. Nesterov and A. Nemirovski. On first-order algorithms for ℓ_1/nuclear norm minimization. *Acta Numerica*, 22:509–575, 2013.

[19] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[20] A. B. Tsybakov. *Introduction to Nonparametric Estimation.* Springer Publishing Company, Incorporated, 1st edition, 2008.

[21] S. A. van de Geer and P. Bhlmann. On the conditions used to prove oracle results for the lasso. *Electron. J. Statist.*, 3:1360–1392, 2009.

A Preliminaries

We start with introducing several objects used in the sequel. We denote $\langle \cdot, \cdot \rangle$ the Hermitian scalar product: for $a, b \in \mathbb{C}^n$, $\langle a, b \rangle = a^H b$. For $a, b \in \mathbb{C}(\mathbb{Z})$ we reserve the shorthand notation

$$\langle a, b \rangle_n := \langle a_0^n, b_0^n \rangle = [a_0^n]^H b_0^n.$$

Convolution matrices. We will extensively use the matrix-vector representation of the discrete convolution.

- Given $y \in \mathbb{C}(\mathbb{Z})$, we associate to it an $(n+1) \times (m+1)$ Toeplitz matrix

$$T(y) = \begin{bmatrix} y_0 & y_{-1} & \dots & y_{-m} \\ y_1 & y_0 & \dots & y_{1-m} \\ \dots & \dots & \dots & \dots \\ y_n & y_{n-1} & \dots & y_{n-m} \end{bmatrix}. \tag{14}$$

  such that $[\varphi * y]_0^n = T(y)\varphi_0^m$ for $\varphi \in \mathbb{C}(\mathbb{Z}_0^m)$. Its squared Frobenius norm satisfies

$$\|T(y)\|_F^2 = \sum_{\tau=0}^{m} \|\Delta^\tau y\|_{n,2}^2. \tag{15}$$

- Given $\varphi \in \mathbb{C}(\mathbb{Z}_0^m)$, consider an $(n+1) \times (m+n+1)$ matrix

$$M(\varphi) = \begin{bmatrix} \varphi_m & \varphi_{m-1} & \dots & \varphi_0 & 0 & 0 & \dots & 0 \\ 0 & \varphi_m & \dots & \dots & \varphi_0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 0 & \varphi_m & \dots & \varphi_0 \end{bmatrix}. \tag{16}$$

  For $y \in \mathbb{C}(\mathbb{Z})$ we have $[\varphi * y]_0^n = M(\varphi) y_{-m}^n$ and

$$\|M(\varphi)\|_F^2 = (n+1)\|\varphi\|_{m,2}^2. \tag{17}$$

- Given $\varphi \in \mathbb{C}(\mathbb{Z}_0^m)$, consider the following circulant matrix of size $m+n+1$:

$$C(\varphi) = \begin{bmatrix} \varphi_m & \dots & \dots & \varphi_0 & 0 & 0 & \dots & \dots & 0 \\ 0 & \varphi_m & \dots & \dots & \varphi_0 & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 0 & \varphi_m & \dots & \dots & \varphi_0 \\ \varphi_0 & 0 & \dots & \dots & \dots & 0 & \varphi_m & \dots & \varphi_1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \varphi_{m-1} & \dots & \varphi_0 & 0 & \dots & \dots & \dots & 0 & \varphi_m \end{bmatrix}. \tag{18}$$

  One has

$$\|C(\varphi)\|_F^2 = (m+n+1)\|\varphi\|_{m,2}^2.$$

  This matrix is useful since $C(\varphi) y_{-m}^n$ encodes the circular convolution of y_{-m}^n and the zero-padded filter φ_0^{m+n} (recall that $\varphi \in \mathbb{C}(\mathbb{Z}_0^m)$) which is diagonalized by the DFT. Specifically,

$$C(\varphi) = F_{m+n}^H D(\varphi) F_{m+n}, \quad \text{where } D(\varphi) = \sqrt{m+n+1}\, \mathrm{diag}(F_{m+n}\, \varphi_0^{m+n}). \tag{19}$$

Deviation bounds. We use the following simple facts about Gaussian random vectors.

- Let $\zeta \sim \mathbb{C}\mathcal{N}(0, I_n)$ be a standard complex Gaussian vector meaning that $\zeta = \xi_1 + \imath\xi_2$ where $\xi_{1,2}$ are two independent draws from $\mathcal{N}(0, I_n)$. We will use a simple bound

$$\text{Prob}\left\{\|\zeta\|_\infty \le \sqrt{2\ln n + 2u}\right\} \ge 1 - e^{-u} \tag{20}$$

  which may be checked by explicitly evaluating the distribution since $|\zeta_1|_2^2 \sim \text{Exp}(1/2)$.

- The following deviation bounds for $\|\zeta\|_2^2 \sim \chi^2_{2n}$ are due to [15, Lemma 1]:

$$\text{Prob}\left\{\frac{\|\zeta\|_2^2}{2} \le n + \sqrt{2nu} + u\right\} \ge 1 - e^{-u}, \quad \text{Prob}\left\{\frac{\|\zeta\|_2^2}{2} \ge n - \sqrt{2nu}\right\} \ge 1 - e^{-u}. \tag{21}$$

  By simple algebra we obtain an upper bound for the norm:

$$\text{Prob}\left\{\|\zeta\|_2 \le \sqrt{2n} + \sqrt{2u}\right\} \ge 1 - e^{-u}. \tag{22}$$

- Further, let K be an $n \times n$ Hermitian matrix with the vector of eigenvalues $\lambda = [\lambda_1; ...; \lambda_n]$. Then the real-valued quadratic form $\zeta^H K\zeta$ has the same distribution as $\xi^T B\xi$, where $\xi = [\xi_1; \xi_2] \sim \mathcal{N}(0, I_{2n})$, and B is a real $2n \times 2n$ symmetric matrix with the vector of eigenvalues $[\lambda; \lambda]$. Hence, we have $\text{Tr}(B) = 2\text{Tr}(K)$, $\|B\|_F^2 = 2\|K\|_F^2$ and $\|B\| = \|K\| \le \|K\|_F$, where $\|\cdot\|$ and $\|\cdot\|_F$ denote the spectral and the Frobenius norm of a matrix. Invoking [15, Lemma 1] again (a close inspection of the proof shows that the assumption of positive semidefiniteness can be relaxed), we have

$$\text{Prob}\left\{\frac{\zeta^H K\zeta}{2} \le \text{Tr}(K) + (u + \sqrt{2u})\|K\|_F\right\} \ge 1 - e^{-u}. \tag{23}$$

  Further, when K is positive semidefinite, we have $\|K\|_F \le \text{Tr}(K)$, whence

$$\text{Prob}\left\{\frac{\zeta^H K\zeta}{2} \le \text{Tr}(K)(1 + \sqrt{u})^2\right\} \ge 1 - e^{-u}. \tag{24}$$

B Proof of Theorems 2.1 and 2.2

B.1 Proof idea

Despite the striking similarity with the Lasso [19], [3], [2], the recoveries of section 2 are of quite different nature. First of all, the ℓ_1-minimization in these methods is aimed to recover a filter but not the signal itself, and this filter is not sparse.[1] The equivalent of "regression matrices" involved in these methods cannot be assumed to satisfy *Restricted Eigenvalue* or *Restricted Isometry* conditions, usually imposed to prove statistical properties of "classical" ℓ_1-recoveries (see e.g. [4], [21], and references therein). Moreover, being constructed from the noisy signal itself, these matrices depend on the noise, what introduces an extra degree of complexity in the analysis of the properties of these estimators. Yet, proofs of Theorem 2.1 and 2.3 rely on some simple ideas and it may be useful to expose these ideas stripped from the technicalities of the complete proof. Given $y \in \mathbb{C}(\mathbb{Z}_{-n}^n)$ let $T(y)$ be the $(n+1) \times (n+1)$ "convolution matrix," as defined in (14) such that for $\varphi \in \mathbb{C}(\mathbb{Z}_0^n)$ $[\varphi * y]_0^n = T(y)\varphi_0^n$. When denoting $f = F_n\varphi$, the optimization problem in (2) can be recast as a "standard" ℓ_1-constrained least-squares problem with respect to f:

$$\min_{f \in \mathbb{C}^{n+1}} \left\{\|y - A_n f\|_2^2 : \ \|f\|_1 \le \overline{\varrho}/\sqrt{n+1}\right\} \tag{25}$$

[1] Unless we consider recovery of signal composed of harmonic oscillations with frequencies on the DFT grid.

where $A_n = T(y)F_n^{-1}$. Observe that $f^o = F_n\varphi^o$ is feasible in (25) so that

$$\|y - A_n\widehat{f}\|_{n,2}^2 \le \|y - A_n f^o\|_{n,2}^2,$$

where $\widehat{f} = F_n\widehat{\varphi}$, so that

$$\begin{aligned}&\|x - A_n\widehat{f}\|_{n,2}^2 - \|x - A_n f^o\|_{n,2}^2 \le 2\sigma\left(\Re\langle\zeta, x - A_n f^o\rangle_n - \Re\langle\zeta, x - A_n\widehat{f}\rangle_n\right)\\ &\le 2\sigma\big|\langle\zeta, A_n(f^o - \widehat{f})\rangle_n\big| \le 2\sigma\|A_n^H\zeta_0^n\|_\infty\|f^o - \widehat{f}\|_1 \le 4\sigma\|A_n^H\zeta_0^n\|_\infty\tfrac{\overline{\varrho}}{\sqrt{n+1}}\end{aligned}$$

In the "classical" situation, where ζ_0^n is independent of A_n (see, e.g., [11]), the norm $\|A_n^H\zeta_0^n\|_\infty$ is bounded by $c_\alpha\sqrt{\ln n}\max_j\|[A_n]_j\|_2 \le c_\alpha\sqrt{n\ln n}\|A_n\|_\infty$ where $\|A\|_\infty = \max_{i,j}|A_{ij}|$ and c_α is a logarithmic in α^{-1} factor. This would rapidly lead to the bound (7) of the theorem. In the case we are interested in, where A_n incorporates observations y_{-n}^n and thus depends on ζ_0^n, curbing the cross term is more involved and requires extra assumptions, e.g. Assumption A.

B.2 Extended formulation

We will prove a simple generalization of the theorem in the case where the length n of "validation sample" may be different from the length m of the adjusted filter. We consider the "skewed" sample

$$y_\tau = x_\tau + \sigma\zeta_\tau \quad -m \le \tau \le n, \tag{26}$$

with $\kappa_{m,n} := \sqrt{\frac{n+1}{m+1}}$. Accordingly, we assume that the regular recovery $\widehat{\varphi} * y$ uses the filter

$$\widehat{\varphi} \in \mathop{\mathrm{Argmin}}_{\varphi\in\mathbb{C}(\mathbb{Z}_0^m)}\left\{\|y - \varphi * y\|_{n,2} : \|\varphi\|_{m,1}^* \le \overline{\varrho}/\sqrt{m+1}\right\}. \tag{27}$$

The corresponding modification of Assumption A is as follows:

Assumption A′ *Let $\mathcal{S}$ be a (unknown) shift-invariant linear subspace of $\mathbb{C}(\mathbb{Z})$, $\Delta\mathcal{S} = \mathcal{S}$, of dimension s, $1 \le s \le n+1$. We suppose that x admits the decomposition:*

$$x = x^{\mathcal{S}} + \varepsilon,$$

where $x^{\mathcal{S}} \in \mathcal{S}$, and ε is "small", namely,

$$\|\Delta^\tau\varepsilon\|_{n,2} \le \sigma\varkappa, \quad 0 \le \tau \le m.$$

In what follows we use the following convenient reformulation of Assumption A′ (reformulation of Assumption A when $m = n$):

> *There exists an s-dimensional (complex) subspace $\mathcal{S}_n \subset \mathbb{C}^{n+1}$ and an idempotent Hermitian $(n+1)\times(n+1)$ matrix $\Pi_{\mathcal{S}_n}$ of rank s – the projector on $\mathcal{S}_n$ – such that*
>
> $$\big\|(I_{n+1} - \Pi_{\mathcal{S}_n})[\Delta^\tau x]_0^n\big\|_2 \left[= \|\Delta^\tau\varepsilon\|_{n,2}\right] \le \sigma\varkappa,\ \tau = 0, ..., m \tag{28}$$
>
> *where I_{n+1} is the $(n+1)\times(n+1)$ identity matrix.*

Theorem B.1. *Let $m, n \in \mathbb{Z}_+$, $\overline{\varrho} \ge 1$, and let $\varphi^o \in \mathbb{C}(\mathbb{Z}_0^m)$ be such that*

$$\|\varphi^o\|_1^* \le \overline{\varrho}/\sqrt{m+1}.$$

Suppose that Assumption A′ holds for some $s \in \mathbb{Z}_+$ and $\varkappa < \infty$. Then for any $\alpha, 0 < \alpha \le 1$, there is a set $\Xi \subset \mathbb{C}^{2n+1}$, $\mathrm{Prob}\{\zeta_{-m}^n \in \Xi\} \ge 1 - \alpha$, of "good realisations" of ζ such that whenever $\zeta_{-m}^n \in \Xi$,

$$\|x - \widehat{\varphi} * y\|_{n,2} \le \|x - \varphi^o * y\|_{n,2} + 2\sigma\left[\sqrt{\overline{\varrho}V_\alpha^2 + (\overline{\varrho}+1)c_\alpha\varkappa} + \sqrt{2s} + c_\alpha\right], \tag{29}$$

where $c_\alpha := \sqrt{2\ln[16/\alpha]}$, and

$$V_\alpha^2 = 2\left(1 + 4\kappa_{m,n}\right)^2\ln\left[55(m+n+1)/\alpha\right].$$

B.3 Proof of Theorem B.1

1^o. The oracle filter φ^o is feasible in (27), hence,

$$\begin{aligned} \|x-\widehat{\varphi}*y\|_{n,2}^2 &= \|(1-\varphi^o)*y\|_{n,2}^2-\sigma^2\|\zeta\|_{n,2}^2-2\,\Re\sigma\langle\zeta,x-\widehat{\varphi}*y\rangle_n \\ &= \|x-\varphi^o*y\|_{n,2}^2-2\,\underbrace{\Re\sigma\langle\zeta,x-\widehat{\varphi}*y\rangle_n}_{\delta^{(1)}}+2\,\underbrace{\Re\sigma\langle\zeta,x-\varphi^o*y\rangle_n}_{\delta^{(2)}}. \end{aligned} \tag{30}$$

Let us bound $\delta^{(1)}$. Denote for brevity $I:=I_{n+1}$, and recall that $\Pi_{\mathcal{S}_n}$ is the projector on $\mathcal{S}_n$ from (28). We have the following decomposition:

$$\begin{aligned} \delta^{(1)} &= \underbrace{\Re\sigma\langle\zeta_0^n,\Pi_{\mathcal{S}_n}[x-\widehat{\varphi}*y]_0^n\rangle}_{\delta_1^{(1)}}+\underbrace{\Re\sigma\langle\zeta_0^n,(I-\Pi_{\mathcal{S}_n})[x-\widehat{\varphi}*x]_0^n\rangle}_{\delta_2^{(1)}} \\ &\quad-\underbrace{\Re\sigma^2\langle\zeta_0^n,(I-\Pi_{\mathcal{S}_n})[\widehat{\varphi}*\zeta]_0^n\rangle}_{\delta_3^{(1)}} \end{aligned} \tag{31}$$

One can easily bound $\delta_1^{(1)}$ under the premise of the theorem:

$$\left|\delta_1^{(1)}\right| \le \sigma\big\|\Pi_{\mathcal{S}_n}\zeta_0^n\big\|_2\big\|\Pi_{\mathcal{S}_n}[x-\widehat{\varphi}*y]_0^n\big\|_2 \le \sigma\big\|\Pi_{\mathcal{S}_n}\zeta_0^n\big\|_2\big\|x-\widehat{\varphi}*y\big\|_{n,2}.$$

Note that $\Pi_{\mathcal{S}_n}\zeta_0^n\sim\mathbb{C}\mathcal{N}(0,I_s)$, and by (22) we have

$$\operatorname{Prob}\left\{\big\|\Pi_{\mathcal{S}_n}\zeta_0^n\big\|_2\ge\sqrt{2s}+\sqrt{2u}\right\}\le e^{-u},$$

obtaining the bound

$$\operatorname{Prob}\left\{|\delta_1^{(1)}|\le\sigma\big\|x-\widehat{\varphi}*y\big\|_{n,2}\left(\sqrt{2s}+\sqrt{2\ln[1/\alpha_1]}\right)\right\}\ge 1-\alpha_1. \tag{32}$$

2^o. We are to bound the second term of (31). To this end, note first that

$$\delta_2^{(1)}=\Re\sigma\langle\zeta_0^n,(I-\Pi_{\mathcal{S}_n})x_0^n\rangle-\Re\sigma\langle\zeta_0^n,(I-\Pi_{\mathcal{S}_n})[\widehat{\varphi}*x]_0^n\rangle.$$

By (28), $\|(I-\Pi_{\mathcal{S}_n})x_0^n\|_2\le\sigma\varkappa$, thus with probability $1-\alpha$,

$$|\langle\zeta_0^n,(I-\Pi_{\mathcal{S}_n})x_0^n\rangle|\le\sigma\varkappa\sqrt{2\ln[1/\alpha]}. \tag{33}$$

On the other hand, using the notation defined in (14), we have $[\widehat{\varphi}*x]_0^n=T(x)\widehat{\varphi}_0^m$, so that

$$\langle\zeta_0^n,(I-\Pi_{\mathcal{S}_n})[\widehat{\varphi}*x]_0^n\rangle=\langle\zeta_0^n,(I-\Pi_{\mathcal{S}_n})T(x)\widehat{\varphi}_0^m\rangle.$$

Note that $[T(x)]_\tau=x_{-\tau}^{-\tau+n}$ for the columns of $T(x)$, $0\le\tau\le m$. By (28), $(I-\Pi_{\mathcal{S}_n})T(x)=T(\varepsilon)$, and by (15),

$$\|(I-\Pi_{\mathcal{S}_n})T(x)\|_F^2=\|T(\varepsilon)\|_F^2=\sum_{\tau=0}^m\big\|\varepsilon_{-\tau}^{-\tau+n}\big\|_2^2\le(m+1)\sigma^2\varkappa^2.$$

Due to (24) we conclude that

$$\big\|T(x)^H(I-\Pi_{\mathcal{S}_n})\zeta_0^n\big\|_2^2\le 2(m+1)\sigma^2\varkappa^2\left(1+\sqrt{\ln[1/\alpha]}\right)^2$$

with probability at least $1-\alpha$. Since

$$|\langle\zeta_0^n,(I-\Pi_{\mathcal{S}_n})T(x)\widehat{\varphi}_0^m\rangle|\le\frac{\overline{\varrho}}{\sqrt{m+1}}\,\big\|T(x)^H(I-\Pi_{\mathcal{S}_n})\zeta_0^n\big\|_2,$$

we arrive at the bound with probability $1-\alpha$:

$$|\langle\zeta_0^n,(I-\Pi_{\mathcal{S}_n})T(x)\widehat{\varphi}_0^m\rangle|\leq\sqrt{2}\sigma\varkappa\overline{\varrho}\left(1+\sqrt{\ln[1/\alpha]}\right).$$

Along with (33) this results in the following bound:

$$\mathrm{Prob}\left\{\left|\delta_2^{(1)}\right|\leq\sqrt{2}\sigma^2\varkappa(\overline{\varrho}+1)\left(1+\sqrt{\ln[2/\alpha_2]}\right)\right\}\geq 1-\alpha_2. \tag{34}$$

3^o. Let us rewrite $\delta_3^{(1)}$ as follows:

$$\delta_3^{(1)}=\Re\sigma^2\langle\zeta_0^n,(I-\Pi_{\mathcal{S}_n})M(\widehat{\varphi})\zeta_{-m}^n\rangle=\Re\sigma^2\langle\zeta_{-m}^n,QM(\widehat{\varphi})\zeta_{-m}^n\rangle,$$

where $M(\widehat{\varphi})\in\mathbb{C}^{(n+1)\times(m+n+1)}$ is defined by (16), and $Q\in\mathbb{C}^{(m+n+1)\times(n+1)}$ is given by

$$Q=\left[\begin{array}{ll}0_{m,n+1}; & I-\Pi_{\mathcal{S}_n}\end{array}\right];$$

hereinafter we denote $0_{m,n}$ the $m\times n$ zero matrix. Now, by the definition of $\widehat{\varphi}$ and since the mapping $\varphi\mapsto M(\varphi)$ is linear,

$$\begin{aligned}\delta_3^{(1)} &= \tfrac{1}{2}(\zeta_{-m}^n)^H\underbrace{(QM(\widehat{\varphi})+M(\widehat{\varphi})^HQ^H)}_{K_1(\widehat{\varphi})}z_{-m}^n\leq\frac{\sigma^2\overline{\varrho}}{2\sqrt{m+1}}\max_{\substack{u\,\in\,\mathbb{C}(\mathbb{Z}_0^m),\\ \|u\|_{m,1}^*\,\leq\,1}}(\zeta_{-m}^n)^HK_1(u)\zeta_{-m}^n\\ &= \frac{\sigma^2\overline{\varrho}}{\sqrt{m+1}}\max_{1\leq j\leq m+1}\max_{\theta\in[0,2\pi]}\tfrac{1}{2}(\zeta_{-m}^n)^HK_1(e^{\imath\theta}u^j)\zeta_{-m}^n,\end{aligned} \tag{35}$$

where $u^j\in\mathbb{C}(\mathbb{Z}_0^m)$, and $[u^j]_0^m=F_m^{-1}\mathrm{e}_j$, e_j being the j-th canonical orth of $\mathbb{R}^{m+1}$. Indeed, $M(\varphi)$ attains its maximum over the convex set

$$\mathcal{B}_{m,1}^*=\{u\in\mathbb{C}(\mathbb{Z}_0^m),\quad\|u\|_{m,1}^*\leq 1\}. \tag{36}$$

at an extremal point $e^{\imath\theta}u_j$, $\theta\in[0,2\pi]$. It is easy to verify that

$$K_1(e^{\imath\theta}u)=K_1(u)\cos\theta+K_2(u)\sin\theta$$

for the Hermitian matrix

$$K_2(u)=\imath\left(QM(u)-M(u)^HQ^H\right).$$

Denoting $q_i^j(\zeta)=\frac{1}{2}(\zeta_{-m}^n)^HK_i(u^j)\zeta_{-m}^n$ for $i=1,2$, we have

$$\begin{aligned}&\max_{\theta\in[0,2\pi]}\tfrac{1}{2}(\zeta_{-m}^n)^HK_1(e^{\imath\theta}u^j)\zeta_{-m}^n=\max_{\theta\in[0,2\pi]}\left[q_1^j(\zeta)\cos\theta+q_2^j(\zeta)\sin\theta\right]\\ &=\sqrt{\left|q_1^j(\zeta)\right|^2+\left|q_2^j(\zeta)\right|^2}\leq\sqrt{2}\max\left(\left|q_1^j(\zeta)\right|,\left|q_2^j(\zeta)\right|\right).\end{aligned} \tag{37}$$

By simple algebra and using (17), we get

$$\mathrm{Tr}\left[K_i(u^j)^2\right]\leq 4\ \mathrm{Tr}[M(u^j)M(u^j)^H]=4(n+1)\|u^j\|_{m,2}^2\leq 4(n+1),\quad i=1,2.$$

Now let us bound $\mathrm{Tr}[K_i(u)]$, $i=1,2$, on the set (36). One may check that for the circulant matrix $C(u)$, cf. (18), it holds:

$$QM(u)=\underbrace{QQ^H}_{R}C(u),$$

where $R=QQ^H$ is an $(m+n+1)\times(m+n+1)$ projection matrix of rank s defined by

$$R=\left[\begin{array}{c|c}0_{m,m} & 0_{m,n+1}\\ \hline 0_{n+1,m} & I-\Pi_{\mathcal{S}_n}\end{array}\right].$$

Hence, denoting $\|\cdot\|_*$ the nuclear norm, we can bound $\mathrm{Tr}[K_i(u)]$, $i = 1, 2$, as follows:

$$\big|\,\mathrm{Tr}[K_i(u)]\big| \le 2\big|\,\mathrm{Tr}[RC(u)]\big| \le 2\|R\|\,\|C(u)\|_* \le 2\|C(u)\|_* = 2\sqrt{m+n+1}\|u\|^*_{m+n,1},$$

where in the last transition we used (19). The following technical lemma gives an upper bound on the norm of a zero padded filter (see Appendix B.4 for the proof):

Lemma B.1. *For any $u \in \mathcal{B}^*_{m,1}$, see (36), and $n \ge 1$, we have*

$$\|u\|^*_{m+n,1} \le \sqrt{1+\kappa^2_{m,n}}(\ln[m+n+1]+3).$$

Thus we arrive at

$$\big|\mathrm{Tr}[K_i(u^j)]\big| \le 2\sqrt{m+1}(\kappa^2_{m,n}+1)(\ln[m+n+1]+3), \quad i = 1, 2.$$

By (23) we conclude that for any fixed pair $(i, j) \in \{1, 2\} \times \{1, ..., m+1\}$, with probability $1-\alpha$,

$$\big|q^j_i(\zeta)\big| \le \big|\mathrm{Tr}[K_i(u^j)]\big| + \big\|K_i(u^j)\big\|_F\left(1+\sqrt{\ln[2/\alpha]}\right)^2.$$

With $\alpha_0 = 2(m+1)\alpha$, by the union bound together with (35) and (37) we get

$$\mathrm{Prob}\left\{\delta^{(1)}_3 \le 2\sqrt{2}\sigma^2\overline{\varrho}\left[(\kappa^2_{m,n}+1)(\ln[m+n+1]+3) + \kappa_{m,n}\left(1+\sqrt{\ln\left[4(m+1)/\alpha_0\right]}\right)^2\right]\right\} \ge 1-\alpha_0. \quad (38)$$

4^o. Bounding $\delta^{(2)}$ is a relatively simple task since φ^o does not depend on the noise. We decompose

$$\delta^{(2)} = \sigma\Re\langle\zeta, x - \varphi^o * x\rangle_n - \sigma^2\Re\langle\zeta, \varphi^o * \zeta\rangle_n.$$

Note that $\Re\langle\zeta, x-\varphi^o * x\rangle_n \sim \mathbb{C}\mathcal{N}(0, \|x-\varphi^o * x\|^2_{n,2})$, therefore, with probability $1-\alpha$,

$$\Re\langle\zeta, x-\varphi^o * x\rangle_n \le \sqrt{2\ln[1/\alpha]}\|x-\varphi^o * x\|_{n,2}. \quad (39)$$

On the other hand,

$$\begin{aligned}\|x-\varphi^o * x\|_{n,2} &\le \|x-\varphi^o * y\|_{n,2} + \sigma\|\varphi^o * \zeta\|_{n,2}\\ &\le \|x-\varphi^o * y\|_{n,2} + \sqrt{2}\sigma\varrho\kappa_{m,n}\left(1+\sqrt{\ln[1/\alpha]}\right)\end{aligned} \quad (40)$$

with probability $1-\alpha$. Indeed, one has

$$\|\varphi^o * \zeta\|^2_{n,2} = \big\|M(\varphi^o)\zeta^n_{-m}\big\|^2_2,$$

where for $M(\varphi^o)$ by (17) we have

$$\|M(\varphi^o)\|^2_F = (n+1)\|\varphi^o\|^2_{m,2} \le \kappa^2_{m,n}\varrho^2. \quad (41)$$

Using (24) we conclude that, with probability at least $1-\alpha$,

$$\|\varphi^o * \zeta\|^2_{n,2} \le 2\kappa^2_{m,n}\varrho^2\left(1+\sqrt{\ln[1/\alpha]}\right)^2,$$

which implies (40). Using (39) and (40), we get that with probability at least $1-\alpha_3$,

$$\begin{aligned}&\Re\langle\zeta, x-\varphi^o * x\rangle_n\\ &\quad\le \sqrt{2\ln\left[2/\alpha_3\right]}\left[\|x-\varphi^o * y\|_{n,2} + \sqrt{2}\sigma\varrho\kappa_{m,n}\left(1+\sqrt{\ln\left[2/\alpha_3\right]}\right)\right]\end{aligned}$$

$$\leq \quad \|x-\varphi^o*y\|_{n,2}\sqrt{2\ln[2/\alpha_3]}+2\sigma\varrho\kappa_{m,n}\left(1+\sqrt{\ln[2/\alpha_3]}\right)^2. \qquad (42)$$

The indefinite quadratic form

$$\Re\langle\zeta,\varphi^o*\zeta\rangle_n=\frac{(\zeta^n_{-m})^H K_0(\varphi^o)\zeta^n_{-m}}{2},$$

where $K_0(\varphi^o)=[0_{m,m+n+1};\,M(\varphi^o)]+[0_{m,m+n+1};\,M(\varphi^o)]^H$, can be bounded similarly to 3^o. We get

$$|\operatorname{Tr}[K_0(\varphi^o)]|\leq 2(n+1)\,|\varphi^o_m|\leq 2\kappa^2_{m,n}\varrho.$$

Indeed, for $e_{m+1}=[0;\,...;\,0;\,1]\in\mathbb{R}^{m+1}$ one has

$$|\varphi^o_m|=|\langle[\varphi^o]^m_0,e_{m+1}\rangle|\leq\|\varphi^o\|^*_{m,1}\|F_m e_{m+1}\|_\infty\leq\frac{\varrho}{m+1}$$

since $\|F_m e_{m+1}\|_\infty=1/\sqrt{m+1}$. By (41), $\|K_0(\varphi^o)\|_F^2\leq 4\,\|M(\varphi^o)\|_F^2\leq 4\kappa^2_{m,n}\varrho^2$. Hence by (23),

$$\operatorname{Prob}\left\{-\Re\langle\zeta,\varphi^o*\zeta\rangle_n\leq 2\kappa^2_{m,n}\varrho+2\kappa_{m,n}\varrho\left(1+\sqrt{2\ln[1/\alpha_4]}\right)^2\right\}\geq 1-\alpha_4. \qquad (43)$$

5^o. It remains to combine the bounds obtained in 1^o-4^o. For any $\alpha\in(0,1]$, putting $\alpha_0=\alpha_1=\alpha_4=\alpha/4$, $\alpha_2=\alpha_3=\alpha/8$, and using the the union bound, we get from (30) with probability $\geq 1-\alpha$:

$$\begin{array}{rcl}
\|x-\widehat{\varphi}*y\|^2_{n,2} & \leq & \|x-\varphi^o*y\|^2_{n,2}+2\delta^{(2)}-2\delta^{(1)}\\
\text{[by (42)]} & \leq & \|x-\varphi^o*y\|^2_{n,2}+2\sigma\|x-\varphi^o*y\|_{n,2}\sqrt{2\ln[16/\alpha]}\\
\text{[by (42), (43)]} & & +4\sigma^2\varrho\Big[\kappa^2_{m,n}+2\kappa_{m,n}\big(1+\sqrt{2\ln[16/\alpha]}\big)^2\Big]\\
\text{[by (32)]} & & +2\sigma\|x-\widehat{\varphi}*y\|_{n,2}\big(\sqrt{2s}+\sqrt{2\ln[16/\alpha]}\big)\\
\text{[by (34)]} & & +2\sqrt{2}\sigma^2(\overline{\varrho}+1)\big(1+\sqrt{\ln[16/\alpha]}\big)\varkappa\\
\text{[by (38)]} & & +4\sqrt{2}\sigma^2\overline{\varrho}\left[(\kappa^2_{m,n}+1)(\ln[m+n+1]+3)+\kappa_{m,n}\left(1+\sqrt{\ln[16(m+1)/\alpha]}\right)^2\right].
\end{array}$$

Hence, denoting

$$\begin{array}{rcl}
\upsilon^{(1)} & = & 2\sqrt{2}\sigma^2(\overline{\varrho}+1)\big(1+\sqrt{\ln[16/\alpha]}\big)\varkappa,\\
\upsilon^{(2)} & = & 4\sqrt{2}\sigma^2\overline{\varrho}\left[(\kappa^2_{m,n}+1)(\ln[m+n+1]+3)+\kappa_{m,n}\left(1+\sqrt{\ln[16(m+1)/\alpha]}\right)^2\right]\\
& & +4\sigma^2\varrho\Big[\kappa^2_{m,n}+2\kappa_{m,n}\big(1+\sqrt{2\ln[16/\alpha]}\big)^2\Big]
\end{array}$$

we obtain

$$\begin{array}{rcl}
\|x-\widehat{\varphi}*y\|^2_{n,2} & \leq & \|x-\varphi^o*y\|^2_{n,2}+2\sigma\big(\sqrt{2s}+\sqrt{2\ln[16/\alpha]}\big)\big(\|x-\widehat{\varphi}*y\|_{n,2}+\|x-\varphi^o*y\|_{n,2}\big)\\
& & +\upsilon^{(1)}+\upsilon^{(2)}.
\end{array}$$

The latter implies that

$$\|x-\widehat{\varphi}*y\|_{n,2}\leq\|x-\varphi^o*y\|_{n,2}+2\sqrt{2}\sigma\big(\sqrt{s}+\sqrt{\ln[16/\alpha]}\big)+\sqrt{\upsilon^{(1)}+\upsilon^{(2)}}.$$

Finally, we arrive at (29) using the bounds

$$\upsilon^{(1)}\leq 4\sqrt{2}\sigma^2(\overline{\varrho}+1)\sqrt{\ln[16/\alpha]}\varkappa.$$

and

$$\begin{array}{l}
\upsilon^{(2)}\leq\sigma^2\overline{\varrho}\left(4\sqrt{2}(\kappa^2_{m,n}+1)(\ln[m+n+1]+4)+4.5(4\sqrt{2}+8)\kappa_{m,n}\ln[16(m+1)/\alpha]\right)\\
\leq 8\sigma^2\overline{\varrho}\,(1+4\kappa_{m,n})^2\ln[55(m+n+1)/\alpha]\,.
\end{array}$$

□

B.4 Proof of Lemma B.1

The function $\|u\|^*_{m+n,1}$ is convex on (36), so its maximum over this set is attained at one the extreme points $F_m[u^j]_0^m = e^{\imath\theta}e_j$ where e_j is the j-th canonical orth of $\mathbb{R}^{m+1}$ and $\theta \in [0, 2\pi]$. Since $u^j_\tau = \frac{1}{\sqrt{m+1}} \exp\left[\imath\theta - \frac{2\pi\imath\tau j}{m+1}\right]$, we obtain

$$\|u^j\|^*_{m+n,1} = \frac{1}{\gamma}\sum_{k=0}^{m+n}\left|\sum_{\tau=0}^{m}\exp\left[\imath\underbrace{2\pi\left(\frac{k}{m+n+1}-\frac{j}{m+1}\right)}_{\omega_{jk}}\tau\right]\right| = \frac{1}{\gamma}\sum_{k=0}^{m+n}\left|D_m\left(\frac{\omega_{jk}}{2}\right)\right|,$$

where $\gamma = \sqrt{(m+n+1)(m+1)}$, and the Dirichlet kernel $D_m(\cdot)$ is defined as

$$D_m(x) := \begin{cases} \frac{\sin((m+1)x)}{\sin(x)}, & x \neq \pi l, \\ m+1, & x = \pi l. \end{cases}$$

Hence, $\gamma\|u^j\|^*_{m+n,1} \le \max_{\epsilon\in[0,\pi]} S_{m+n}(\epsilon)$, where

$$S_{m,n}(\epsilon) = \sum_{k=0}^{m+n}\left|D_m\left(\frac{\pi k}{m+n+1} - \epsilon\right)\right|, \quad \epsilon \in [0,\pi]. \tag{44}$$

Note that $|D_m(x)|$ is upper bounded by the following (positive) function on the circle $\mathbb{R}/\pi\mathbb{Z}$:

$$B_m(x) = \begin{cases} \frac{\pi}{2\min(x,\pi-x)}, & x \in (0,\pi), \\ m+1, & x = 0. \end{cases}$$

For any $\epsilon \in [0,\pi]$, the summation in (44) is over a regular $(m+n+1)$-grid on $\mathbb{R}/\pi\mathbb{Z}$. The contribution to the sum of each of two closest to zero points of the grid is at most $\max_{x\in[0,\pi]} D_m(x) = m+1$. For the remaining points, we can upper bound $|D_m(x)| \le B_m(x)$ noting that $B_m(x)$ decreases over $[\frac{\pi}{m+n+1}, \frac{\pi}{2}]$ as long as $n \ge 1$. These considerations result in

$$S_{m,n}(\epsilon) \le 2(m+1) + \sum_{k=1}^{\lceil\frac{m+n-1}{2}\rceil}\frac{m+n+1}{k} \le 2(m+1) + (m+n+1)\left(\ln\left\lceil\frac{m+n+1}{2}\right\rceil + 1\right)$$

where in the last transition we used the simple bound $H_n \le \ln n + 1$ for harmonic numbers. □

B.5 Proof of Theorem 2.2

We decompose

$$\begin{aligned} |[x - \widehat{\varphi} * y]_n| &= |[(\phi^o + (1-\phi^o)) * (x - \widehat{\varphi} * y)]_n| \\ &\le \quad |[\phi^o * (x - \widehat{\varphi} * y)]_n| + |[(1-\widehat{\varphi}) * (1-\phi^o) * x]_n| + \sigma|[\widehat{\varphi} * \zeta]_n| + \sigma|[\widehat{\varphi} * \phi^o * \zeta]_n| \\ &:= \quad \delta^{(1)} + \delta^{(2)} + \delta^{(3)} + \delta^{(4)}. \end{aligned} \tag{45}$$

We have

$$\delta^{(1)} \le \|\phi^o\|_2 \|x - \widehat{\varphi} * y)\|_{m,2} \le \frac{\rho}{\sqrt{m+1}}\|x - \widehat{\varphi} * y\|_{m,2}.$$

When using the bound of Corollary 2.1 with $\overline{\varrho} = 2\rho^2$, we conclude that, with probability $\ge 1 - \alpha/3$,

$$\delta^{(1)} \le c\frac{\sigma\rho}{\sqrt{n}}\left[\rho^2\sqrt{\ln[1/\alpha]} + \rho\sqrt{\left(\varkappa \ln\left[1/\alpha\right] + \ln\left[n/\alpha\right]\right)} + \sqrt{s}\right].$$

Next we get

$$\delta^{(2)} \leq (1+\|\widehat{\varphi}\|_1)\,\|(1-\phi^o)*x\|_{n,\infty} \leq (1+2\rho^2)\frac{\sigma\rho}{\sqrt{m+1}}.$$

By the Parseval identity,

$$\delta^{(3)} = \sigma|\langle F_n[\widehat{\varphi}^*]_0^n, F_n[\zeta]_{0,n}\rangle| \leq \sigma\|\widehat{\varphi}\|^*_{n,1}\|\zeta\|^*_{n,\infty} \leq \frac{2\sigma\rho^2}{\sqrt{m+1}}\sigma\sqrt{2\ln[3(n+1)/\alpha]},$$

where the last inequality, holding with probability $\geq 1-\alpha/3$, is due to (20).

Finally, observe that, with probability $\geq 1-\alpha/3$ (cf. (46)),

$$\|\phi^o*\zeta\|_{n,2} \leq \sqrt{2}\rho\left(1+\sqrt{\ln[3/\alpha]}\right).$$

Therefore, we have for $\delta^{(4)}$:

$$\delta^{(4)} \leq \sigma\|\widehat{\varphi}\|_{n,2}\|\phi^o*\zeta\|_{n,2} \leq \sigma\frac{2\rho^2}{\sqrt{m+1}}\sqrt{2}\rho\left(1+\sqrt{\ln[3/\alpha]}\right) = \sigma\frac{2\sqrt{2}\rho^3}{\sqrt{m+1}}\left(1+\sqrt{\ln[3/\alpha]}\right)$$

with probability $1-\alpha/3$. When substituting the bound for $\delta^{(k)}$, $k=1,...,4$, into (45) we arrive at the result of the theorem. $\square$

C Miscellaneous proofs

Proof of relations (4) **and** (5)**.** Let $n=2m$, $\phi\in\mathbb{C}(\mathbb{Z}_0^m)$, and let $\varphi\in\mathbb{C}(\mathbb{Z}_0^n)$ satisfy $\varphi=\phi*\phi$. Then

$$\begin{aligned}\|\varphi\|^*_{n,1} &= (2m+1)^{-1/2}\sum_{k=0}^{2m}|[F_{2m}\varphi_0^{2m}]_k| = \sqrt{2m+1}\sum_{k=0}^{2m}\left(\frac{|[F_{2m}\phi_0^{2m}]_k|}{\sqrt{2m+1}}\right)^2\\ &= \sqrt{2m+1}\|\phi\|^{*2}_{2m,2} = \sqrt{2m+1}\|\phi\|^2_{2m,2} = \sqrt{2m+1}\|\phi\|^2_{m,2} \leq \frac{\sqrt{2m+1}\,\rho^2}{m+1},\end{aligned}$$

implying (4). Moreover, since $1-\phi*\phi=(1+\phi)*(1-\phi)$, for all $x\in\mathbb{C}(\mathbb{Z})$ one has for all $\tau\in\mathbb{Z}$:

$$\begin{aligned}|x_\tau-[\varphi^o*x]_\tau| &= |[(1+\phi^o)*(1-\phi^o)*x]_\tau| = \left|\sum_{j=0}^m[1+\phi^o]_j\,[x-\phi^o*x]_{\tau-j}\right|\\ &\leq \|1+\phi^o\|_1\max_{0\leq j\leq m}|[x-\phi^o*x]_{\tau-j}| \leq \frac{\sigma(1+\rho)\rho}{\sqrt{m+1}}\end{aligned}$$

(we have used (3) to obtain the last inequality), and

$$\|x-[\varphi^o*x]\|_{n,2} \leq \sigma(1+\rho)\rho.$$

Next note that

$$\|\varphi^o*\zeta\|^2_{n,2} = \langle\zeta, M(\varphi^o)\zeta\rangle_n,$$

where $M(\varphi)$ is defined as in (16). When taking into account that

$$\|\varphi\|_2 \leq \|\varphi\|^*_{n,1} \leq \frac{2\rho^2\sqrt{n+1}}{n+2},$$

we get (cf. (17)) $\|M(\varphi^o)\|_F^2 = (n+1)\|\varphi\|_2^2 \leq 4\rho^4$, so that the concentration inequality (24) now implies that, given $0<\alpha\leq 1$, with probability at least $1-\alpha$,

$$\|\varphi^o*\zeta\|_{n,2} \leq 2\sqrt{2}\rho^2\left(1+\sqrt{\ln[1/\alpha]}\right), \tag{46}$$

and we arrive at (5).

Proof of (9). Assume that Assumption A holds true for some $n \geq s$ and $\varepsilon \equiv 0$. Let $\Pi_{\mathcal{S}_n}$ be the Euclidean projector on the space $\mathcal{S}_n$ of elements of $\mathcal{S}$ restricted on $\mathbb{C}(\mathbb{Z}_0^n)$. Since $\dim(\mathcal{S}_n) \leq s$, $\|\Pi_{\mathcal{S}_n}\|_2^2 = \mathrm{Tr}(\Pi_{\mathcal{S}_n}) \leq s$, there is $\iota \in \{0, ..., n\}$ such that the ι-th column $[\Pi_{\mathcal{S}_n}]_\iota$ of $\Pi_{\mathcal{S}_n}$ satisfies $\|[\Pi_{\mathcal{S}_n}]_\iota\|_2 \leq \sqrt{s/(n+1)}$. Note that one has $x_\iota - \langle [\Pi_{\mathcal{S}_n}]_\iota, x_0^n \rangle = 0$, and $\mathcal{S}_n$ is time-invariant, implying that

$$x_\tau - \langle [\Pi_{\mathcal{S}_n}]_\iota, x_{\tau-\iota}^{\tau-\iota+n} \rangle = 0, \ \ \forall \tau \in \mathbb{Z}.$$

We conclude that there is $\phi^o \in \mathbb{C}(\mathbb{Z}_{-n}^n)$, $\phi^o = [0; ...; 0; [\Pi_{\mathcal{S}_n}]_\iota; 0; ...; 0]$ (i.e., vector $[\Pi_{\mathcal{S}_n}]_\iota$ completed with zeros in such a way that ι-th element of $[\Pi_{\mathcal{S}_n}]_\iota$ becomes the central $(n+1-$th) entry of ϕ^o) such that

$$\|\phi^o\|_2 \leq \sqrt{s/(n+1)}, \ \text{ and } \ x_\tau - [\phi^o * x]_\tau = 0, \ \ \forall \tau \in \mathbb{Z}. \qquad \square$$

C.1 Proof of Theorem 2.4

Note that to prove the lemma we have to exhibit a vector $q \in \mathbb{C}^{m+1}$ of small ℓ_2-norm and such that the polynomial $1 - q(z) = 1 - \left[\sum_{i=0}^m q_i z^i\right]$ is divisible by $p(z)$, i.e., that there is a polynomial $r(z)$ of degree $m - s$ such that

$$1 - q(z) = r(z)p(z).$$

Indeed, this would imply that

$$x_t - [q * x]_t = [1 - q(\Delta)]x_t = r(\Delta)p(\Delta)x_t = 0$$

due to $p(\Delta)x_t = 0$,

The bound $\|q\|_2 \leq C' s^{3/2}\sqrt{\frac{\ln s}{m}}$ of (11) is proved in [14, Lemma 6.1]. Our objective is to prove the "remaining" inequality

$$\|q\|_2 \leq C' s \sqrt{\frac{\ln[ms]}{m}}.$$

So, let $\theta_1, ..., \theta_s$ be complex numbers of modulus 1 – the roots of the polynomial $p(z)$. Given $\delta = 1 - \epsilon \in (0, 1)$, let us set $\bar{\delta} = 2\delta/(1+\delta)$, so that

$$\frac{\bar{\delta}}{\delta} - 1 = 1 - \bar{\delta} > 0. \tag{47}$$

Consider the function

$$\bar{q}(z) = \prod_{i=1}^s \frac{z - \theta_i}{\delta z - \theta_i}.$$

Note that $\bar{q}(\cdot)$ has no singularities in the circle

$$\mathcal{B} = \{z : |z| \leq 1/\bar{\delta}\};$$

besides this, we have $\bar{q}(0) = 1$.Let $|z| = 1/\bar{\delta}$, so that $z = \bar{\delta}^{-1} w$ with $|w| = 1$. We have

$$\frac{|z - \theta_i|}{|\delta z - \theta_i|} = \frac{1}{\delta} \frac{|w - \bar{\delta}\theta_i|}{|w - \frac{\bar{\delta}}{\delta}\theta_i|}.$$

We claim that when $|w| = 1$, $|w - \bar{\delta}\theta_i| \leq |w - \frac{\bar{\delta}}{\delta}\theta_i|$.

> Indeed, assuming w.l.o.g. that w is not proportional to θ_i, consider triangle Δ with the vertices $A = w$, $B = \bar{\delta}\theta_i$ and $C = \frac{\bar{\delta}}{\delta}\theta_i$. Let also $D = \theta_i$. By (47), the segment $\overline{AD}$ is a median in Δ, and $\angle CDA$ is $\geq \frac{\pi}{2}$ (since D is the closest to C point in the unit circle, and the latter contains A), so that $|w - \bar{\delta}\theta_i| \leq |w - \frac{\bar{\delta}}{\delta}\theta_i|$.

As a consequence, we get

$$z \in \mathcal{B} \;\Rightarrow\; |\bar{q}(z)| \leq \delta^{-s}, \tag{48}$$

whence also

$$|z| = 1 \;\Rightarrow\; |\bar{q}(z)| \leq \delta^{-s}. \tag{49}$$

Now, the polynomial $p(z) = \prod_{i=1}^{s}(z - \theta_i)$ on the boundary of $\mathcal{B}$ clearly satisfies

$$|p(z)| \geq \left[\frac{1}{\bar{\bar{\delta}}} - 1\right]^s = \left[\frac{1-\delta}{2\delta}\right]^s,$$

which combines with (48) to imply that the modulus of the holomorphic in $\mathcal{B}$ function

$$\bar{r}(z) = \left[\prod_{i=1}^{s}(\delta z - \theta_i)\right]^{-1}$$

is bounded with $\delta^{-s}\left[\frac{1-\delta}{2\delta}\right]^{-s} = \left[\frac{2}{1-\delta}\right]^s$ on the boundary of $\mathcal{B}$. It follows that the coefficients r_j of the Taylor series of $\bar{r}$ satisfy

$$|r_j| \leq \left[\frac{2}{1-\delta}\right]^s \bar{\delta}^j, \quad j = 0, 1, 2, \ldots$$

When setting

$$q^\ell(z) = p(z) r^\ell(z), \quad r^\ell(z) = \sum_{j=1}^{\ell} r_j z^j, \tag{50}$$

for $|z| \leq 1$, utilizing the trivial upper bound $|p(z)| \leq 2^s$, we get

$$\begin{array}{rcl} |q^\ell(z) - \bar{q}(z)| & = & |p(z)|[r^\ell(z) - \bar{r}(z)]| \leq |p(z)||r^\ell(z) - \bar{r}(z)| \\ & \leq & 2^s \left[\frac{2}{1-\delta}\right]^s \sum_{j=\ell+1}^{\infty} |r_j| \leq \left[\frac{4}{1-\delta}\right]^s \frac{\bar{\delta}^{\ell+1}}{1-\bar{\delta}}. \end{array} \tag{51}$$

Note that $q^\ell(0) = p(0)r^\ell(0) = p(0)\bar{r}(0) = 1$, that q^ℓ is a polynomial of degree $\ell + s$, and that q^ℓ is divisible by $p(z)$. Besides this, on the unit circumference we have, by (51),

$$|q^\ell(z)| \leq |\bar{q}(z)| + \left[\frac{4}{1-\delta}\right]^s \frac{\bar{\delta}^{\ell+1}}{1-\bar{\delta}} \leq \delta^{-s} + \underbrace{\left[\frac{4}{1-\delta}\right]^d \frac{\bar{\delta}^{\ell+1}}{1-\bar{\delta}}}_{R} \tag{52}$$

(we have used (49)). Now,

$$\bar{\delta} = \frac{2\delta}{1+\delta} = \frac{2-2\epsilon}{2-\epsilon} = \frac{1-\epsilon}{1-\epsilon/2} \leq 1 - \epsilon/2 \leq \mathrm{e}^{-\epsilon/2},$$

and

$$\frac{1}{1-\bar{\delta}} = \frac{1+\delta}{1-\delta} = \frac{2-\epsilon}{\epsilon} \leq \frac{2}{\epsilon}.$$

We can upper-bound R:

$$R = \left[\frac{4}{1-\delta}\right]^s \frac{\bar{\delta}^{\ell+1}}{1-\bar{\delta}} \leq \frac{2^{2s+1}}{\epsilon^{s+1}} \mathrm{e}^{-\epsilon\ell/2}$$

Now, given positive integer ℓ and positive α such that

$$\frac{\alpha}{\ell} \leq \tfrac{1}{4}, \tag{53}$$

let $\epsilon = \frac{\alpha}{2\ell s}$. Since $0 < \epsilon \leq \frac{1}{8}$, we have $-\ln(\delta) = -\ln(1-\epsilon) \leq 2\epsilon = \frac{\alpha}{\ell s}$, implying that $\bar{\delta} \leq \mathrm{e}^{-\epsilon/2} = \mathrm{e}^{-\frac{\alpha}{4\ell s}}$, and

$$R \leq \left[\frac{8\ell s}{\alpha}\right]^{s+1} \exp\{-\frac{\alpha}{4s}\}.$$

Now let us put

$$\alpha = \alpha(\ell, s) = 4s(s+2)\ln(8\ell s);$$

observe that this choice of α satisfies (53), provided that

$$\ell \geq O(1)s^2 \ln(s+1) \tag{54}$$

with properly selected absolute constant $O(1)$. With this selection of α, we have $\alpha \geq 1$, whence

$$\begin{aligned} R\left[\frac{\alpha}{\ell}\right]^{-1} &\leq \exp\left\{-\frac{\alpha}{4s}\right\}\left[\frac{8\ell s}{\alpha}\right]^{s+1}\frac{\ell}{\alpha} \leq \exp\left\{-\frac{\alpha}{4s}\right\}[8\ell s]^{s+2} \\ &\leq \exp\{-(s+2)\ln(8\ell s)\}\exp\{(s+2)\ln(8\ell s)\} = 1, \end{aligned}$$

that is,

$$R \leq \frac{\alpha}{\ell} \leq \tfrac{1}{4}. \tag{55}$$

Furthermore,

$$\begin{aligned} \delta^{-s} &= \exp\{-s\ln(1-\epsilon)\} \leq \exp\{2\epsilon s\} = \exp\{\tfrac{\alpha}{\ell}\} \leq 2, \\ \delta^{-2s} &= \exp\{-2s\ln(1-\epsilon)\} \leq \exp\{4\epsilon s\} = \exp\{\tfrac{2\alpha}{\ell}\} \leq 1 + \exp\{\tfrac{1}{2}\}\tfrac{2\alpha}{\ell} \leq 1 + \tfrac{4\alpha}{\ell}. \end{aligned} \tag{56}$$

When invoking (52) and utilizing (56) and (55) we get

$$\frac{1}{2\pi}\oint_{|z|=1} |q^\ell(z)|^2 |dz| \leq \delta^{-2s} + 2\delta^{-s}R + R^2 \leq 1 + 4\frac{\alpha}{\ell} + 4R + \frac{1}{4}R \leq 1 + 10\frac{\alpha}{\ell}.$$

On the other hand, denoting by q_0, q_1,...,$q_{\ell+s}$ the coefficients of the polynomial q^ℓ and taking into account that $\bar{q}_0 = q^\ell(0) = 1$, we have

$$1 + \sum_{i=1}^{\ell+s} |q_i|^2 = |q_0|^2 + ... + |q_{\ell+s}|^2 = \frac{1}{2\pi}\oint_{|z|=1} |q^\ell(z)|^2 |dz| \leq 1 + 10\frac{\alpha}{\ell}. \tag{57}$$

We are done: when denoting $m = \ell + s$, and $q(z) = \sum_{i=1}^m q_j z^j$, we have the vector of coefficients $q = [0; q_1; ...; q_m] \in \mathbb{C}^{m+1}$ of $q(z)$ such that, by (57),

$$\|q\|_2^2 \leq \frac{40s(s+2)\ln[8s(m-s)]}{m-s},$$

and such that the polynomial $q^\ell(z) = 1 + q(z)$ is divisible by $p(z)$ due to (50). □

D Comments on algorithm implementation

In Sec. 2 we considered only recoveries which estimated the value x_t of the signal via the observations at $n+1$ points $t-n, ..., t$ "on the left" (filtering problem). To recover the whole signal, one may consider a more flexible alternative – *interpolating* recovery – which estimates x_t using observations on the left *and* on the right of t. In particular, if the objective is to recover a signal on the interval $\{-n, ..., n\}$, one can apply interpolating recoveries using the same observations $y_{-n}, ..., y_n$ to estimate x_τ for each $\tau \in \{-n, ..., n\}$.

Ideally, when using pointwise recovery, a specific filter is constructed for each time instant t. This may pose a tremendous amount of computation, for instance, when recovering a high-resolution image.

Alternatively, one may split the signal into blocks, and process the points of each block using the same filter (cf. e.g. Theorem 2.1). For instance, a one-dimensional signal can be divided into blocks of length, say, $2m+1$, and to recover $x \in \mathbb{C}(\mathbb{Z}_{-m}^{m})$ in each block one may fit one filter of length $m+1$ recovering the right "half-block" x_0^m and another filter recovering the left "half-block" x_{-m}^{-1}.

When recovering a signal or an image, the "optimal" filter bandwidth – the length of the filter for which the balance of the stochastic and the approximation error is attained – depends on the local signal parameters (e.g. smoothness) and vary from one point to another. In the experiments of Sec. 3, we implement pointwise and blockwise bandwidth adaptation using a procedure similar in spirit to the celebrated Lepski adaptation procedure [17]. Details are given in e.g. [13, Sec. 3.3] and [7, Sec. 3.2].

Note that the optimization problems (2) and (8) underlying recovery algorithms are well structured Second-Order Conic Programs (SOCP) and can be solved using Interior-point methods (IPM). However, the computational complexity of IPM applied to SOCP with *dense* matrices grows rapidly with problem dimension, so that large problems of this type arising in signal and image processing are well beyond the reach of these techniques. On the other hand, these problems possess nice geometry associated with complex ℓ_1-norm. Moreover, their *first-order information* – the value of objective and its gradient at a given φ – can be computed using Fast Fourier Transform is time which is almost linear in problem size. Therefore, we used first-order optimization algorithms, such as Mirror-Prox (MP) and Nesterov's accelerated gradient algorithms (see [18] and reference therein), in our recovery implementation. A complete description of the application of these optimization algorithms to our problems is beyond the scope of the paper; we shall present it elsewhere.